DOCUMENT RESUME

ED 110 467 95 TM 004 737

AUTHOR Nassif, Paula M.
TITLE Systematic, Cooperative Evaluation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE May 75
NOTE 21p.; Paper presented at the New England Educational Research Organization Conference (Provincetown, Massachusetts, May 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Curriculum Evaluation; Elementary Secondary Education; *Evaluation Methods; *Mathematics Curriculum; Mathematics Instruction; Measurement Techniques; *Program Evaluation; *Success Factors
IDENTIFIERS Elementary Secondary Education Act Title III; ESEA Title III

ABSTRACT

Evaluation procedures based on a systematic evaluation methodology, decision-maker validity, new measurement and design techniques, low cost, and a high level of cooperation on the part of the school staff were used in the assessment of a public school mathematics program for grades 3-8. The mathematics curriculum was organized into Spirals which approximated grade levels but allowed students to pass from one Spiral to another within grades. The evaluation staff consisted of members from the school program: the superintendent, the budget analyst, and the math-science supervisor; members from the funding organization: the director of the Regional Education Council and the evaluation coordinator; and the evaluators: the program evaluator and the evaluation consultant. Evaluation success was attributed to the insurances for preserving decision-maker validity, the cooperation from the school system, and the enhanced data interpretations due to matrix sampling techniques. (Author/BJG)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED110467

SYSTEMATIC, COOPERATIVE EVALUATION

Paula M. Nassif
Student Affairs Research & Evaluation Office
245 Whitmore Administration Building
University of Massachusetts
Amherst, Massachusetts 01002

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 004 737

A paper presented at the
New England Educational Research Organization Conference
Provincetown, Massachusetts
May 1975

Introduction

The evaluation conducted for a public school mathematics program, Grades 3-8, is an example of thorough and useful evaluation procedures based on: (1) a systematic evaluation methodology, (2) decision-maker validity, (3) new measurement and design techniques, (4) low cost, and (5) a high level of cooperation on the part of the school staff.

The success of the evaluation is due in large part to its being systematic and cooperative. It was systematic because the evaluation was conducted using an organized, operational, standardized set of rules and procedures designed to provide data for decision-making. The evaluation was cooperative because the largest possible number of teachers and staff members were involved for the purpose of obtaining their input at particular steps in the evaluation. In so doing, many more evaluation activities were accomplished than internal (within the school system) and external (organizational funding) resources would originally have allowed. For example, all students were tested, using consumable tests, and each student was correctly assigned to a test measuring him at his current performance level.

The focus of this paper is to describe the school mathematics curriculum and to present the steps of the evaluation, an explanation of the mode of implementation of each step, and the ways in which cooperation and systemization enhanced the evaluation.

Curriculum

The mathematics curriculum was organized into Spirals which approximated grade levels but allowed students to pass from one Spiral to another within grades. At the beginning of the evaluation there were six Spirals, each composed of a curriculum guide which listed broad attitudinal goals and a comprehensive statement of cognitive objectives developed for the content covered by the Spiral. At that time, there was not a complete set of test items designed to measure the cognitive objectives. There were no attitudinal items.

Within each classroom, a variety of instructional materials were provided to assist students in acquiring the skills stated in the objectives. Most classes were individualized to the extent that students progressed freely from one Spiral to the next by demonstrating mastery on a prepared Spiral curriculum test. Therefore, within each classroom several Spirals were generally represented.

Methodology

As stated earlier, a methodology is an organized, operational, standardized set of rules and procedures designed to accomplish a specified task. A model, on the other hand, is not operational; it may provide some general principles or guidelines but does not fully explain the procedures to be undertaken. In that regard, a methodology primarily has utility through being: (1) more flexible than fixed models (which are often applicable only within a limited range of

convenience); (2) prescriptive rather than descriptive (providing procedures rather than principles); and (3) dynamic (providing for evaluation and revision contingent upon application).

An adaptation of the Fortune/Hutchinson Evaluation Methodology (Benedict, 1973) was used for this program. This evaluation methodology has as its purpose: "to provide data for decision-making." Therefore, an implication of the purpose is that the data provided must be used in decision-making for the evaluation to be considered successful (Hutchinson, 1972). The level of correctness or appropriateness of the data provided to the decision-maker is referred to as decision-maker validity. Decision-maker validity is measured by the amount of data provided by the evaluation which is used by the decision-maker in making decisions.

Therefore, a high level of decision-maker validity was necessary to assess the evaluation as worthwhile. Procedures were used at each step of the evaluation to insure decision-maker validity of the data. These procedures will be explained throughout the remainder of this paper.

The methodology included the following 12 major steps:

STEP 1. The Enterprise--that aspect of the instructional situation (such as an individual student, a group, or the curriculum) about which decisions are to be made and, therefore, about which data are to be collected--is defined by the person initiating the evaluation.

STEP 2. The Resources available for the evaluation are specified by the person initiating the evaluation.

STEP 3. The Decision-Maker, who makes decisions about the enterprise and to whom data are to be reported, is identified by the person initiating the evaluation. (More than one decision-maker may be chosen. The decision-maker may be an individual, a group, or a collection of people. The initiator may or may not be selected as the evaluator.)

STEP 4. The Goals of the enterprise are elicited from the decision-maker by the evaluator but not stated by the evaluator.

STEP 5. The Objectives associated with each goal are elicited from the decision-maker by the evaluator but not stated by the evaluator.

STEP 6. The Decisions to be made about each objective are elicited from the decision-maker by the evaluator but not stated by the evaluator.

STEP 7. The Measurement Technique (or items) is developed by the evaluator or the decision-maker and is criterion-referenced--i.e., related to an individual objective.

STEP 8. The Design, consisting of "which measurement instrument to be administered to which subjects on what schedule," is created by the evaluator and approved by the decision-maker. The design usually includes multiple matrix sampling.

STEP 9. The Collection of the data is carried out by systematic logistics procedures.

STEP 10. The Analysis of the data collected is performed quickly and flexibly by computer to the specifications set by the evaluator and approved by the decision-maker.

STEP 11. The Report of the data is printed by computer and submitted to the specific decision-maker for whom the evaluation was designed.

STEP 12. The Revision of the evaluation is considered by the decision-maker.

Each of these major steps has several operationally defined sub-steps.

Initiating the Evaluation

The evaluation staff consisted of members from the school program, i.e. the Superintendent, the Budget Analyst, and the Math-Science Supervisor; members from the funding organization, i.e. the Director of the Regional Education Council and the Evaluation Coordinator; and the evaluators, i.e. the Program Evaluator and the Evaluation Consultant.

Steps 1, 2 and 3 of the Evaluation Methodology are considered initiating the evaluation; these steps were accomplished at the first meeting of the evaluation staff and school personnel. The purpose of Step 1, defining the enterprise, was to state explicitly the parameters of the program which would be included in the evaluation. Who or what would be evaluated, when and where the evaluation would take place were included in the identification of the enterprise. This step was accomplished by the use of a prepared questionnaire which was administered to the school personnel attending the first meeting with the evaluation staff. The questionnaire listed several possible parameters of the enterprise, such as math texts, staff, math audio-visual equipment, and math affective and cognitive skills. These parameters were rank ordered by the attending school personnel according to "importance" to them. The number of parameters to be evaluated was selected considering the amount of available resources. The results of the questionnaire

indicated that the enterprise to be evaluated was the mathematics program, Grades 3-8, cognitive and affective domains.

The purpose of Step 2, determining and allocating resources, was to provide a flexible estimate of the amount of time to be spent in implementing each evaluation task. The total amount of available resources (money, time) was determined by contractual arrangement.

Presenting the resource allocations to the school personnel was accomplished by the use of a prepared chart which matched the outline of potential evaluation activities to a distribution of resources. The distribution of the evaluator's time was based on documentation in the Fortune/Hutchinson Evaluation Methodology of the approximate percentage of total resources required for each step.

The third step in initiating the evaluation was identifying the decision-makers to whom data would be provided. This step was also accomplished by means of a prepared questionnaire. The position titles of several people associated with the mathematics curriculum who were potential decision-makers were rank ordered by the school personnel in attendance at the meeting. The list included people such as the Superintendent, Math-Science Supervisor, school board and parents. The rank ordering was based on several criteria that specified that the person(s) chosen as decision-maker should have sufficient knowledge of the math program, time to devote to evaluation procedures, rapport with teachers and administrators, and commitment to use the evaluation information provided. Since the number of decision-makers for whom evaluation data will be provided depends largely on the amount of available resources, it was decided by the school personnel and regional staff that one

primary decision-maker would be selected: the district-wide Math-Science Supervisor. This indicated that the evaluator was primarily responsible for providing data for the decision-making needs of the Math-Science Supervisor. In addition, the math curriculum team would consult on correctness of math skills and appropriateness of procedures within the classroom at various steps throughout the evaluation.

Articulating Goals (Step 4)

The process for articulating goals used in this evaluation was shortened from the methodology because goal statements of the math program had been previously generated. Using these prepared statements, the goals were isolated (broken down) by the evaluator so that each statement contained one goal intent. These statements were randomly arranged on a questionnaire and prepared for rank ordering by the decision-maker. The goals were ranked in priority order by the decision-maker according to the criterion that goals, when measured, would provide the most "important" information for decision-making. The following goal was selected as the focus of the evaluation:

> To provide for each student the sequential math program that will work toward personalization so that each student has continuous progress in his own understanding of math.

Specifying Objectives (Step 5)

Specification of objectives involved many procedures undertaken by several groups of people. To begin with, the goal stated above had to be defined in all its component parts. Each of the components of a goal is an objective. The task of specifying objectives was begun by

the decision-maker with the assistance of the evaluator.

Cognitive objectives. The "sequential math program" as stated in the goal was defined by the objectives in the Spiral Curriculum Guides prepared for each grade level by the math teachers. Therefore, the task of specifying cognitive objectives became one of arranging the objectives in priority order according to selected criteria and then determining the maximum number of objectives possible for testing. In this way, each goal was presented in its operational components by a statement of its related objectives. Furthermore, the objectives were arranged in priority order according to criteria determined by the decision-maker. These criteria were (1) that the objectives involve skills that the students would use in the future and (2) that the objectives be content continuous within Spirals and across Spirals. The math curriculum team was consulted at this point. Along with the decision-maker, each member of the team ranked the Spiral Guide objectives according to the criteria stated above. This task was completed by the entire curriculum team, thereby providing the evaluation staff with much information. A weighting system was used to coordinate the ranking results and arrive at a final rank order for each Spiral level.

Affective objectives. The phrase stating that the math program would "work toward personalization . . ." became the basis for the affective mathematics attitude measurement. The process of "Operationalization of Fuzzy Concepts" (Hutchinson & Benedict, 1970) was used to specify the affective objectives from the affective goal ("work toward

personalization . . ."). The decision-maker performed the "Operationalization" and established 10 characteristics (performance indicators) of a student who had personalized his math program. (Since this procedure is lengthy, it will not be presented in this paper. The reader is asked to refer to the actual document.) The characteristics included resourcefulness, ability to see the relevance of math outside the classroom, and creativity. The math curriculum team also worked through the process of "Operationalization." However, team members operationally defined each of the 10 characteristics generated by the decision-maker from the term "personalization of the math curriculum." The purpose of this procedure was twofold. Primarily, each affective characteristic had to be defined to as nearly behavioral a level as possible for measurement purposes. Secondarily, input from the curriculum team suggested issues regarding actual classroom activities or problem areas which should be treated with some sensitivity.

The teachers also provided a check that vocabulary and language level did not exceed the students' reading abilities at lower grade levels. A final selection of cognitive and affective objectives was made by the decision-maker.

The manner in which this step (specifying objectives) was carried out was important in that the original establishment of decision-maker validity from selection of goals was preserved while delineating both cognitive and affective objectives. As a result, data provided about these cognitive and affective components would meet decision-making needs and contribute to decision-maker validity. In addition, this

step saw the first substantial involvement of the curriculum team, whose cooperation throughout the evaluation enhanced its success.

Delineating Decisions (Step 6)

The process of delineating decisions was conducted by the evaluator and the decision-maker. The purpose of this step was to anticipate potential decisions to be made about the enterprise at the completion of the evaluation. The decision-maker was asked to project the type and content of decisions that would most assist him in determining changes needed for improving the enterprise.

After a list of these potential decisions had been identified and arranged in priority order by the decision-maker, the decisions were examined for correspondence with at least one objective. If an objective did not match at least one decision, it was examined with regard to need in providing pertinent, usable data to the decision-maker. If it was determined that the objective was necessary, then a corresponding decision was added. If a decision did not have a matching objective from which data would be provided, either an objective was added or the decision was eliminated because it was judged unnecessary.

This act of matching objectives to potential decisions was a further extension of the insurance of maintaining decision-maker validity throughout the evaluation. As a result, at this point no unnecessary objectives, which would have produced irrelevant data, were maintained and no anticipated decision was without a matching objective.

Measurement Technique (Step 7)

Cognitive testing. Determining the measurement technique was the step that required the greatest amount of the evaluator's time. Since resources were limited, it was decided that items previously generated by other school systems would be used for cognitive testing. The following parameters were determined for cognitive testing:

1. Six Spiral tests would be created.
2. Four parallel forms of each test would be developed.
3. Each test would contain 20 items: 15 items pertaining to that Spiral, 3 items from the following Spiral to measure pre-instruction knowledge, and 2 items from the previous Spiral to measure retention.
4. All items would be four-option multiple choice questions.
5. Each student would be individually assigned to the appropriate Spiral test.
6. All students would be tested.
7. All testing and scoring would be criterion-referenced.

The combination of the first three parameters listed above introduces the concept of matrix sampling that was used in this evaluation. This matrix sampling is an integral part of Comprehensive Achievement Monitoring (CAM), a classroom management and assessment system (Gorth, et al., 1975).

A particularly valuable aspect of the testing design was stated in Parameter #3 above. According to the aspects of the CAM system

implemented in this evaluation, student achievement is monitored for various levels of instruction. That is, information is provided on the students' knowledge of content related to objectives for which they have recently received instruction. In addition, information is provided about knowledge of content related to objectives for which they have not received instruction. This "pre-test" information is valuable for making curriculum decisions based on amount of pre-instructional knowledge of material. Finally, information is provided about retention level of content. Again, decisions regarding possible restructuring of the curriculum to provide for more or less review possibilities can be made based on empirical evidence.

According to the CAM system, matrix sampling provides more information about the performance level of the group. Since results were not reported by individual student, it was appropriate to provide group data. For item sampling purposes, four parallel items were selected to measure each objective. The items were coded by number, randomly assigned to a matrix, and selected from the matrix when compiling the four parallel test forms for each Spiral. This method allows for more accurate results and better decision-making, because results of performance on four items measuring each objective are provided.

The items were examined for correctness, continuity, and language level by the decision-maker. After the decision-maker indicated content satisfaction with the items, they were prepared on 5" x 8" banking cards to allow for easy storage and retrieval and multiple use. In this way, tests of any length, measuring any skills, could be immediately constructed by merely pulling the relevant coded item cards from the file

and lining the cards upside down on the surface of a duplicating machine. Cards could be reordered, substituted, etc., to create a multitude of tests for various purposes. No retyping or item generation was required.

The level of cooperation provided by the teachers in the schools allowed the evaluation to measure each student with appropriate criterion-referenced assessment. The teachers were provided with a copy of the objectives that were to be measured for each Spiral. They were asked to indicate for each of their students the highest Spiral which the student had completed. In this way, students took Spiral tests only on material for which they had received instruction. Teachers responded for 100% of the students in Grades 3-8. It is only due to their complete assistance that it was possible not only to measure all students but to assign them to the appropriate criterion-referenced measure.

Affective testing. With regard to the affective measures, two parallel written attitude surveys of 10 questions each were developed to encompass manifestations of the 10 affective characteristics and objectives previously generated. For most of the items, a four-option Likert scale provided the response choices. Students were instructed that there was no "right" or "wrong" answer to these questions. Students in Grades 3-8 were randomly assigned to one of the two forms of the attitude survey.

Designing the Data Collection (Step 8)

The instruments for measuring math cognitive skills and math attitudes were printed on consumable tests. Since all students would be tested, there was no random sampling of students in that regard.

However, since this evaluation involved a matrix sampling design, students were randomly assigned to one of four parallel test forms for their appropriate Spiral level.

Since there was only one evaluator and limited resources, the cooperation of teachers in testing procedures was of importance. The evaluator met with all teachers prior to testing to explain rules for standardizing testing procedures and to answer questions. Before the testing period, each teacher was asked to fill in the requested demographic information on a separate sheet for each student. This sheet would eventually be returned to the teacher and student and would serve as the test answer sheet. After the demographic data for each student were collected, the sheets were returned to the evaluator. At that point the appropriate Spiral test and attitude survey were attached to each student's answer sheet. Teachers were instructed on how to monitor the testing situation. Therefore, on the day assigned for test administration, teachers had only to distribute the package with the answer sheet and tests to the appropriate student.

Data Collection (Step 9)

Since the data collection strategies had been operationally defined and thoroughly prepared, actual test administration was accomplished with relative ease. Data for the evaluation were collected on specified days at all schools during a specified one-hour interval. Each teacher was provided with answer keys for all Spirals, so that if he desired he could derive a performance score for each of his students. By contractual agreement, results of performance labeled by classroom, by teacher

or by individual student would not be available from the computer analysis. Correcting individual student answer sheets is another task that could not have been completed by the evaluator or the decision-maker because of time and resource constraints. Three thousand twenty-five students were tested.

Analyzing Data (Step 10)

Data analysis provided information on the percentage of students who responded to each option of each item for all forms of all tests. These percentages were adjusted for missing values. The correct response was indicated so that percentages could be compared with various incorrect response options. SPSS Codebook provided frequency of response and percentages of students in all grades who responded to a particular Spiral test. In addition, data was CROSSTABBED by grade. In other words, performance on a Spiral test was broken down by performance by students in the various grades assigned to the Spiral level. Data were not analyzed by classroom, student or school building.

With regard to the attitude survey, frequencies of response indicating levels of positive or negative attitude toward the 10 affective objectives were presented in addition to crosstabs, by grade and Spiral.

Reporting Results (Step 11)

Throughout the course of the evaluation, reports were prepared by the evaluator and submitted for acceptance to the decision-maker. The reports contained summaries of each meeting, explanations of steps in the methodology, complete descriptions of steps in the implementation of the evaluation, and all materials distributed or used in work with

students, the curriculum team, or the decision-maker. Therefore, all procedures were fully and satisfactorily documented at the completion of the evaluation.

The results of the evaluation were presented in four forms in the Final Report to the decision-maker. The data were verbally explained for each variable in each form of the cognitive Spiral tests and the attitude survey. The results of each attitude variable were graphed by Spiral and by grade for extreme positive response on each item scale.

The results of all variables in each Spiral test were presented in charts showing differences in performances from students in various grades.

Since matrix sampling was employed, group performance could meaningfully be discussed. The Final Report presented the data and also provided some interpretations in light of curriculum design and development.

Finally, the completely detailed, labeled computer output was included with the report.

Revising the Evaluation (Step 12)

The purpose of the revision of the evaluation was to adapt the methodology described for future use. Since a follow-up evaluation was not immediately conducted, the revision has not been necessary to this point.

If a follow-up evaluation were to assume some or all of the goals, decisions and objectives of the present evaluation, the tasks of the

evaluation would be reduced significantly. The materials for testing could be generated simply from the item banking cards, and all procedures have been sufficiently documented that an additional implicit purpose would be served. That is, the evaluation system could become self-sustaining and able to be conducted eventually without the assistance of external evaluators. Much effort in preparing materials and meeting with the school staff was directed toward training them in the methods involved in evaluation to suit their needs.

Summary/Educational Importance

The success of the implementation of this evaluation is due primarily to cooperation from the decision-maker and the math program teachers, and to a strict maintenance to preserving decision-maker validity. The operational definition of the goal to be measured was determined and approved by the decision-maker himself. Hence, no outside biases were introduced at this point. In addition, the objectives, items and decisions were subjected to additional review against criteria established by the decision-maker. These steps, in addition to those built into the methodology, insured decision-maker validity throughout.

The matrix sampling provided results of group performance and curriculum effectiveness. Hence, decisions could correctly be made regarding these factors.

Finally, the cooperation of the decision-maker and the program teachers provided for measuring all students in the program rather than

a sample of students.

With regard to low cost, many evaluation activities were conducted that resources would not otherwise have allowed. They were completed due to cooperation, as indicated, and to the efficient selection and use of pre-existing, effective materials such as the bank of cognitive items.

The implementation of the methodology is considered successful if the purpose (to provide data for decision-making) is met. The success of the evaluation can thus be attributed to the insurances for preserving decision-maker validity, the cooperation from the school system, and the enhanced data interpretations due to matrix sampling techniques.

This evaluation was partially funded by Project Evaluation, an ESEA Title III project directed by Dr. Philip Saif of the Capitol Region Education Council.

References

Benedict, L. G. The Fortune/Hutchinson evaluation methodology: A decision-oriented approach. A paper presented at the meeting of the American Educational Research Association, New Orleans, 1973.

CO-OP. Catalog of behavioral objectives and criterion-referenced test items. Amherst, Mass.: Author, University of Massachusetts, School of Education, 1972.

Gorth, W. P., O'Reilly, R. P., & Pinsky, P. D. Comprehensive Achievement Monitoring. Englewood Cliffs, N.J.: Educational Technology Publications, 1975.

Hutchinson, T. E. Some overlooked implications of the purpose: To provide data for decision making. A paper presented at the meeting of the American Educational Research Association, Chicago, 1972.

Hutchinson, T. E., & Benedict, L. G. The operationalization of fuzzy concepts. University of Massachusetts. Xerox, 1970.